# Neural network approach for Compton-scattering imaging

**Jiajun Wang and Yuanmei Wang**

*Department of Life Science and Biomedical Engineering, Zhejiang University, Hangzhou 310027, China*

**Zheru Chi**

*Department of Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

The problem of image reconstruction with Compton-scattering spectral data is an ill-posed problem, and the measurement error may be seriously amplified in the reconstruction result. For a stable solution, some kinds of *a priori* models of the problem should be incorporated into the process of reconstruction. Lee *et al.* [IEEE. Trans. Nucl. Sci. **40**, 2049 (1993)] have proposed a continuous model with binary line processes. Owing to the coexistence of the continuous variable and the binary variable, the commonly used optimization methods for problems with continuous variables cannot be used in this case, and therefore a coupled-gradient artificial neural network was proposed for this mixed-integer problem. By introducing two interacting parts (with one part for the continuous variable and the other for the binary line processes) into the network, and by defining the appropriate energy function and dynamics, high-quality solutions were obtained upon convergence of the dynamics. Some simulated results are presented. © 1998 Optical Society of America [S0740-3232(98)00609-7]

*OCIS codes:* 290.0290, 200.4260, 170.0170, 100.3020.

## 1. INTRODUCTION

Since the use of Compton-scattered radiation of high-energy Gamma rays in the examination of the internal electron density in tissues was proposed by Lale in 1959,[1] much theoretical and experimental work has been done in this field. The most important advantage of the Compton-scattering imaging technique lies in the fact that the detected photons contain precise and direct position information and can produce an image directly from the electron density distribution of the tested object. As compared with the transmission tomography technique, in which the opposite-side source–detector geometry is needed, this method can offer flexibility in positioning the detector, which does not have to be on the opposite side of the source. On account of this superiority, the Compton-scattering imaging technique is especially suitable for the noninvasive testing of extended industrial objects or in cases when the object is embedded underground. For superficial testing,[2] the Compton-scattering imaging technique can have relatively high sensitivity. Furthermore, the dose needed for this technique is much smaller than that needed in transmission tomography, an important consideration, especially for medical purposes.

There are different modes from which projecting data can be obtained: the point-by-point mode,[3,4] the line-by-line mode,[5,6] and the plane-by-plane mode.[7] The latter two imaging modes have significantly decreased the radiation dose needed in the imaging process. Farmer and Collins[5] were the first to propose the use of the energy spectrum of scattered gamma ray photons for reconstruction purposes. In this paper we will follow the idea of Farmer and Collins, and the spectral data will be obtained from a plane-by-plane scanning mode (Fig. 1), by computer simulation.

Because of the exponential nonlinear dependence of the attenuation factors on the unknown electron densities, the problem is ill-posed. The ill-posedness of this problem forces the solution to be highly susceptible to noisy or erroneous measurements. To obtain a stable solution for such an ill-posed problem, one should incorporate some kinds of *a priori* knowledge into the reconstruction process. In this paper, we adapt a continuous prior model with binary line processes proposed in Ref. 8 that has been employed in emission tomography. Owing to the coexistence of both binary and continuous variables in this model, a deterministic annealing method has been developed for this mixed-integer optimization problem,[8] in which the solutions were obtained after iterations at a series of decreasing temperatures. In this paper we propose to solve this problem in another way, by the neural optimization method. By introducing two mutually interactive parts into the neural network (with one corresponding to the continuous variable and the other to the binary line processes) and by defining an appropriate energy function and dynamics for the neural network, we obtained high-quality solutions upon convergence of the dynamics. Several simulation results will be presented.

## 2. FORWARD MODEL

When an object such as the tissues of the human body is irradiated by photons with energy higher than 100 KeV (as shown in Fig. 1), Compton scattering will be the dominant mode of interaction between photons and the outer-
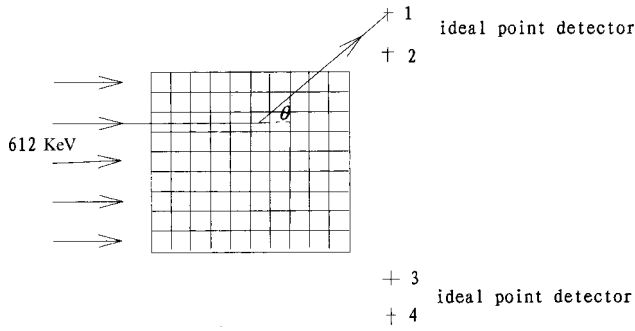
Fig. 1.   Experimental configuration of Compton-scattering imaging.

shell electrons of the target object.   By shielding both the source and the detectors well and after doing some compensation for the multiple-scattering contributions, one may assume that the detector response results from the single-scattered portion.   In this case, when a selected two-dimensional target slice is irradiated by photons with energy $E_0$, the photon flux at the $i$th energy bin $E_i$ of the ideal point detector can be expressed as

$$\psi(E_i) = \sum_{j=1}^{N^2} a_{ij}\rho_j$$

$$= \frac{\Phi}{2\pi} \sum_{j=1}^{N^2} \delta_{ij} f_1(E_0) f_2(E_i) \frac{d\sigma}{d\Omega} (E_0, E_i)\rho_j \Delta V_j / R_j^2, \quad (1)$$

where M and $N^2$ are, respectively, the number of energy bins the ideal point detector can discriminate and the number of voxels included in our image; $\delta_{ij}$ equals 1 if the $j$th voxel is on the $i$th equiscatter line (photons scattered at points on this line have the same energy $E$) or otherwise equals zero and $\Delta V_j$ is the volume of the $j$th voxel; $\rho_j$ denotes the electron density of the $j$th voxel; $(d\sigma/d\Omega) \times (E_0, E)$ is the famous Klein–Nishina differential scattering cross section[9]; $\Phi$ is the incident photon flux; $f_1$ and $f_2$ are the prescattering and postscattering attenuation factors that depend on the unknown electron density; $1/(2\pi)$ accounts for the azimuthal direction where isotropic scattering is assumed; and $R_j^2$ takes into account the photon divergence after being scattered.   The relationship between $E_0$ and $E_i$ is presented in Ref. 10.   When Compton scattering dominates, the prescattering and postscattering attenuation factors can be expressed in terms of the integrated scattering cross section and the unknown electron density as

$$f_1 = \exp\left[-\sum_j \rho_j \sigma_c(E_0)l_j\right],$$

$$f_2 = \exp\left[-\sum_j \rho_j \sigma_c(E_i)l_j\right]. \quad (2)$$

The summation in the first and second expressions of Eqs. (2) are along the incident and the scattering directions, respectively, and $\sigma_c(E_0)$, $\sigma(E_i)$ are the integrated Compton-scattering cross section at energies $E_0$ and $E_i$, which can be integrated from the Klein–Nishina differen-

tial cross section with respect to steradians.   Equation (1) is the so-called forward model.

If the discretized detector response and the voxel values (electron density) are regarded as vectors, the forward model of Eq. (1) can be rewritten in a more compact matrix form,

$$\psi = \mathbf{A}(\rho)\rho, \quad (3)$$

where $\mathbf{A}$ is a matrix whose elements are defined as in Eq. (1).   Owing to the dependence of the attenuation factors on the electron density, the attenuation factors will decay rapidly with the distance that photons travel in the tested object.   Therefore contributions to the detector response from different voxels on the tested sections will be very different, which leads to the ill-conditioning of matrix $\mathbf{A}$ and the ill-posedness of the problem.[11]   For an ill-posed problem, there exists a rapidly decaying singular-value spectrum in matrix $\mathbf{A}$.   (The corresponding condition number, defined as the ratio between the largest and the smallest singular values of the projecting matrix $\mathbf{A}$, reaches the order $10^2$ in our reconstruction problem and will grow rapidly with the scale of the image considered.)   This phenomenon can be clearly seen when we adjust the experimental geometry and the discrimination ability of detectors such that matrix $\mathbf{A}$ becomes diagonal.   However, in general cases, the ill-posedness can be analyzed by the procedure of singular-value decomposition.   As a result of the ill-conditioning of matrix $\mathbf{A}$, a small error in the measurements can be highly amplified in solutions corresponding to small singular values.   To obtain stable solutions, some kinds of prior information should be used to constrain the solution space.   One convenient method of incorporating priors into the inverse problem is the Bayesian method, which will be used in our reconstruction work.

## 3.   RECONSTRUCTION ALGORITHM

We begin our reconstruction work with the Bayesian *a posteriori* probability, given the measurement data or, equivalently, the logarithm of it, i.e.,

$$\ln P(\rho|\psi) = \ln P(\psi|\rho) + \ln P(\rho) - \ln P(\psi). \quad (4)$$

The last term in Eq. (4) can be dropped out since it is a constant and has nothing to do with the electron density.

Since the dominant noise source in our model of functional imaging is photon noise, we use the well-known fact that the detector photon noise is Poisson noise.   Hence the likelihood distribution can be written as

$$P(\psi|\rho) = \prod_{i=1}^{M} \frac{(\bar{\psi}_i)^{\psi_i} \exp(-\bar{\psi}_i)}{\psi_i!}, \quad (5)$$

where $\bar{\psi}_i = \Sigma_{j=1}^{N^2} a_{ij}\rho_j$, with $\psi_i$ being the detector response of the $i$th energy bin.

The second term in Eq. (4) is the prior term through which the *a priori* model can be easily incorporated.   The *a priori* model used here is a continuous functional of the Gibbs form with binary line processes, which can incorporate interactions while allowing abrupt changes across edges or region boundaries[12,13]:

$$P(\rho) = \frac{1}{Z} \exp[-E_P(\rho, l)], \qquad (6)$$

where $Z$ is a normalized partition function, while

$$E_P(\rho, l) = \lambda \sum_j [\rho_v{}^2(j)(1 - l_j{}^h) + \alpha l_j{}^h]$$

$$+ \lambda \sum_j [\rho_h{}^2(j)(1 - l_j{}^v) + \alpha l_j{}^v] \qquad (7)$$

where $\rho_v$ and $\rho_h$ are the difference between neighboring pixels in the vertical and the horizontal directions, respectively, $l^h$ and $l^v$ are the binary-valued $\{0, 1\}$ line processes first proposed by Geman and Geman,[14] and $\lambda$ is a weighting factor determining the extent of smoothing. Here $\alpha$ is a threshold parameter for whether or not the difference operator in Eq. (9) will take an action. A value $l^h = 1$ corresponds to the presence of a horizontal line process. A similar argument can be obtained for $l^v$. $E_P$ is a smoothing function with line processes, which was proposed by Lee *et al.*[8]

Following the ideas of the Bayesian method, the reconstructed solution can be obtained by maximizing the Bayesian *a posteriori* probability or the logarithm of it. From Eqs. (4)–(7), the reconstruction problem can be written in the following form:

$$\rho = \underset{(\rho \geq 0, l)}{\arg \min}\{E(\rho, l) = E_D(\rho) + E_P(\rho, l)\}, \qquad (8)$$

where

$$E_D(\rho) = -\ln P(\psi|\rho) = \sum_{i=1}^{M} [\bar{\psi}_i - \psi_i \ln(\bar{\psi}_i)]$$

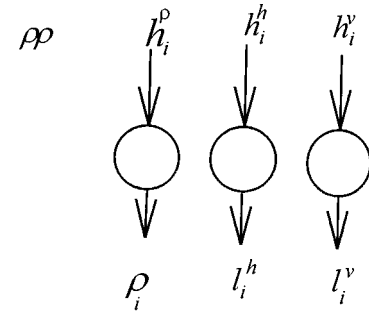$$+ \text{ terms independent of } \rho. \quad (9)$$

Because of the introduction of the binary line processes into the objective function, optimization becomes a problem with mixed binary and continuous variables that is very difficult to solve. In this paper we propose a coupled-gradient artificial neural network for this problem and use a neural optimization method to obtain the solution.

The proposed neural network consists of two mutually interactive recurrent parts, with one corresponding to the continuous variable and the other corresponding to the binary variable. That is, the output of the nodes of the two networks are the electron density $\rho$ (the continuous variable) and the binary line processes $l^h$, $l^v$ (binary variables). The architecture of the neural network is illustrated in Fig. 2.
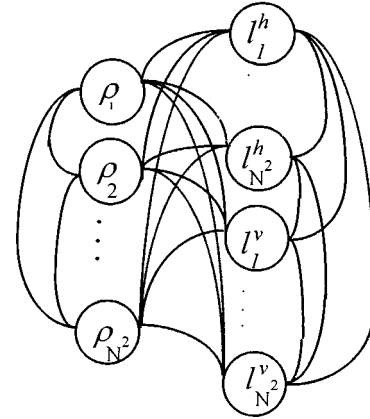
If the domain of $\rho$ is $[0, \rho_M]$ (where $\rho_M$ is the largest electron density of the transverse cross section considered), the state space of the system can be given by

$$\Omega = \prod_{i=1}^{N^2} [0, \rho_M] \times \prod_{i=1}^{N^2} \{0, 1\} \times \prod_{i=1}^{N^2} \{0, 1\}. \qquad (10)$$

The input to the nodes of the two networks as depicted in Fig. 2 will be denoted by the real variables $h_i{}^\rho, h_i{}^h, h_i{}^v$. In order to describe the dynamics of the network by means of differential equations, we relax the domain of the binary line processes $l^h, l^v$ to $[0, 1]$. By introducing



(a)

(b)

Fig. 2. Structure of the coupled-gradient neural network: (a) input and output structure, (b) coupled structure.

some appropriate penalty terms into the energy function of the networks, we can expect that the outputs of nodes of the binary network will be forced to 0 or 1 upon convergence.

In order to guarantee the output $\rho_j \in [0, \rho_M]$, $l_j \in [0, 1]$, we select the activation functions as

$$\rho_j = f_j{}^\rho(h_j{}^\rho) = \frac{\rho_M}{1 + \exp(-h_j{}^\rho)}, \qquad (11)$$

$$l_j{}^x = f_j{}^l(h_j{}^x) = \frac{1}{1 + \exp(-h_j{}^x)}, \qquad x = h \text{ or } v. \qquad (12)$$

The energy function of the neural network includes penalty terms in addition to the objective function of the reconstruction problem:

$$\epsilon(\rho, l) = E(\rho, l) + \beta\left[\sum_{j=1}^{N^2} l_j{}^h(1 - l_j{}^h)\right.$$

$$\left. + \sum_{j=1}^{N^2} l_j{}^v(1 - l_j{}^v)\right], \qquad (13)$$

where $\beta$ is a positive penalty coefficient. It is obvious that the terms in the parentheses of Eq. (13) reach their minimum when $l_j{}^h$, $l_j{}^v = 0$ or 1, which forces the stable outputs of the binary neural network to have binary val-

ues of 0 or 1. Correspondingly, the dynamics of the system can be expressed as the negative gradient of the energy function:

$$h_j{}^{\rho} = -\eta_{\rho} \frac{\partial \epsilon(\rho, l)}{\partial \rho_j}, \qquad \rho_j = f_j{}^{\rho}(h_j{}^{\rho}),$$

$$h_j{}^x = -\eta_l \frac{\partial \epsilon(\rho, l)}{\partial l_j{}^x}, \qquad l_j{}^x = f_j{}^l(h_j{}^x), \qquad x = h \ \text{or} \ v,$$

(14)

where $\eta_{\rho}$, $\eta_l$ are two positive scaling factors between the two interactive networks that determine the evolution dynamics of the network and have nothing to do with the extent of smoothing.

In order to study the stability of the network, let us inspect the evolution process of the energy function with time:

$$\frac{d\epsilon}{dt} = \sum_{j=1}^{N^2} \frac{\partial \epsilon}{\partial \rho_j} \rho_j + \sum_{j=1}^{N^2} \frac{\partial \epsilon}{\partial l_j{}^h} l_j{}^h + \sum_{j=1}^{N^2} \frac{\partial \epsilon}{\partial l_j{}^v} l_j{}^v$$

$$= -\sum_{j=1}^{N^2} \eta_{\rho}{}^{-1} (h_j{}^{\rho})^2 [f_j{}^{\rho}(h_j{}^{\rho})]'$$

$$- \sum_{j=1}^{N^2} \eta_l{}^{-1} (h_j{}^h)^2 [f_j{}^l(h_j{}^h)]'$$

$$- \sum_{j=1}^{N^2} \eta_l{}^{-1} (h_j{}^v)^2 [f_j{}^l(h_j{}^v)]' \leqslant 0. \qquad (15)$$

Hence $\epsilon$ is a Lyapunov function of our neural network, and the network is stable; i.e., the energy is decreasing with time. However, owing to the coexistence of the objective function for the reconstruction problem and the penalty term in the energy function of the network, there is the problem of selecting the parameter $\beta$ that determines the extent of optimization of the objective function and the extent of saturation of the output of the binary network. In this paper, this selection of $\beta$ will be determined by experimental trials.

After the introduction of the neural networks for the reconstruction problem, the only task remaining for us to solve is the dynamic differential equation (15). We will use the forward Euler numerical integral method with the initial point at the center of the state space. Upon convergence of the dynamics of the network, the outputs of the binary network may not be fully 0 or 1, and we need a threshold of 0.5 to obtain the optimal binary values of the line processes. After the determination of the binary line processes, we substitute them for the continuous variable (the electron density) in the dynamic equations and continue to solve them until convergence, after which the results of the reconstruction are obtained.

## 4. COMPUTER SIMULATION RESULTS AND DISCUSSION

In order to examine the performance of our algorithm for Compton-scattering imaging, we used the algorithm for a simulated phantom. The phantom contains two hot regions with pixel values of 200 and 150 and one cold region

of 80; the background pixel value is 100. The projecting data are obtained from Eq. (1), and Poisson-type noise is added. In our simulated experiment, four ideal point detectors are used, which are located symmetrically on one side of the test section. The energy resolution of each ideal detector is 1.5 KeV. A wide beam of parallel gamma ray photons with energy of 662 KeV from a $^{137}$Cs monoenergetic source irradiates the test section. For simplicity, the contamination of the detector response resulting from the multiple-scattered photons and nonscattered photons will not be considered here.

We have used this method for the reconstruction of our simulated phantom with respect to several selections of the weighting parameter $\lambda$. The other unknown coefficients involved in the dynamic equation are determined by a series of trials. Through several trials, we found that $\eta_{\rho} = 100$, $\eta_l = 1$, $\beta = 4$, $\alpha = 5$ constitute a set of good parameters.

To determine the image quality quantitatively, we define the normalized root-mean-square (NRMS) error of the reconstructed image as

$$\text{NRMS} = \left[ \frac{\displaystyle\sum_{j=1}^{N^2} (\rho_j - \hat{\rho}_j)^2}{\displaystyle\sum_{j=1}^{N^2} (\rho_j - \bar{\rho})^2} \right]^{1/2},$$

where

$$\bar{\rho} = \frac{1}{N^2} \sum_{j=1}^{N^2} \hat{\rho}_j, \qquad (16)$$

where $\hat{\rho}_j$ being the pixel value of the simulated phantom and $N^2$ being the total number of pixels in the image.

Figure 3 shows the simulated phantom. Figures 4 and 5 show results corresponding to $\lambda = 0.005$ and $\lambda = 0.1$, respectively; their NRMS values are 0.0705 and 0.6643, respectively. It can be easily seen that the reconstruction results are strongly affected by the weighting factor $\lambda$ and that the results shown in Fig. 5 are overly smoothed. Therefore we should ensure that we are not obtaining a stable solution at the expense of an inaccurate or physically unrealistic solution. For comparison, we also present the results obtained by setting $\lambda = 0$ (Fig. 6), the NRMS of which is 0.9319. Comparing Figs. 4 and 6, we can see that introducing a line-process prior and properly selecting the regularization parameter (the weighting factor) can significantly improve the quality of Compton-scattering imaging.

## 5. SUMMARY

In summary, we used an artificial neural network for the reconstruction of the Compton-scattering image with binary line processes as priors. The quality of the reconstructed image was significantly improved relative to that without priors. Although forward-scattering imaging is studied here, all methods presented are suitable for the backscattering case. However, difficulties still remain. The first is how to determine the penalty parameter (which is common in the Hopfield-type optimization
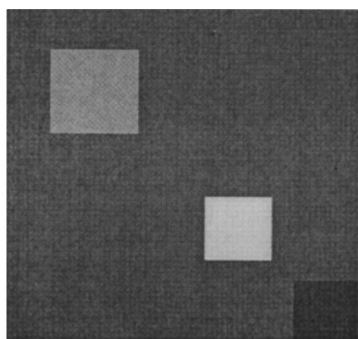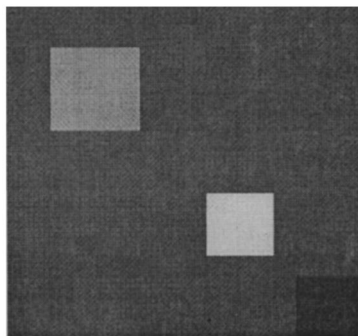
Fig. 3. Computer-simulated phantom.



Fig. 5. Reconstructed result under the condition $\lambda = 0.1$, NRMS = 0.6643.



Fig. 4. Reconstructed result under the condition $\lambda = 0.005$, NRMS = 0.0705.



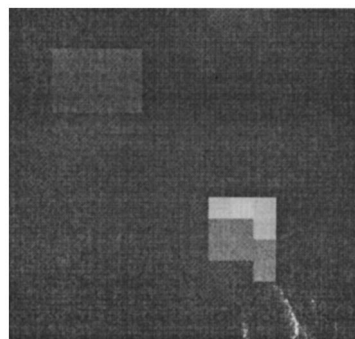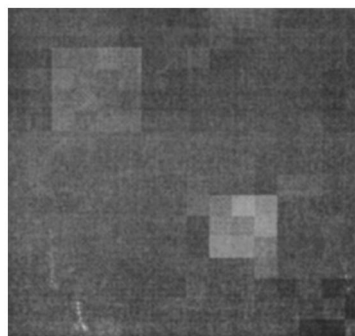Fig. 6. Reconstructed results without priors: $\lambda = 0$, NRMS = 0.9319.

approaches[15]) and the scaled factor involved in the dynamic equations. The second is how to estimate the parameter $\lambda$, which in some way reflects the noise content in the measurements as well as the inherent ill-posedness of the problem. To the knowledge of the authors, at least at present there are no established theories for solving these problems. One rough method for solving this problem is to generalize parameters obtained by running trials for known objects to those for unknown objects. How to estimate those parameters theoretically will be the subject of future research.

## ACKNOWLEDGMENT

## REFERENCES

1. P. G. Lale, "The examination of internal tissues, using gamma ray scatter with a possible extension to megavoltage radiography," Phys. Med. Biol. **4**, 159–166 (1959).
2. G. Harding, "On the sensitivity and application possibilities of a novel Compton scatter imaging," IEEE Trans. Nucl. Sci. **29**, 1260–1265 (1982).
3. J. J. Battista and M. J. Bronskill, "Compton scatter imaging of transverse section: corrections for multiple scattering and attenuation," Phys. Med. Biol. **22**, 229–244 (1977).
4. J. R. Larmarsh, *Introduction to Nuclear Engineering* (Addison-Wesley, Reading, Mass., 1983).
5. F. T. Farmer and M. P. Collins, "A new approach to the determination of anatomical cross sections of the body by Compton scattering of gamma rays," Phys. Med. Biol. **16**, 577 (1970).
6. F. T. Farmer and M. P. Collins, "A further appraisal of the Compton scattering method for determining anatomical cross-sections of the body," Phys. Med. Biol. **19**, 808 (1974).
7. R. Guzzardi, M. Mey, S. Solfanelli, and C. Giuntini, "The 90° Compton scattering tomography of the lung" J. Nucl. Meth. All. Sci. **22**, 11, 1978.
8. M. Lee, A. Rangarajan, I. G. Zubal, and G. Gindi, "A continuation method for emission tomography," IEEE Trans. Nucl. Sci. **40**, 2049–2058 (1993).
9. R. Cesareo, A. L. Hanson, G. E. Gigante, L. J. Pedraza, and S. Q. G. Mahtaboally, "Interaction of Key Photons with matter and new applications," Phys. Rep. **213**, 117–178 (1992).
10. R. Guzzardi and G. Licitra, "A critical review of Compton imaging," CRC Rep. **15**, 237–268 (1988).
11. N. V. Arendtsz and E. M. A. Hussein, "Energy spectral Compton scatter imaging—part 1: theory and mathematics," IEEE Trans. Nucl. Sci. **42**, 2155–2165 (1995).
12. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," IEEE. Trans. Med. Imaging **8**, 194–202 (1989).
13. J. Besag, "On the statistical analysis of dirty pictures," J. R. Statist. Soc. B **48**, 259–302 (1986).
14. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," IEEE Trans. Patt. Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
15. G. Wilson and G. Pawley, "On the stability of the traveling salesman problem algorithm of Hopfield and Tank," Biol. Cybern. **58**, 63–70 (1988).